#9: Getting A/B Testing Right

In the previous lesson we talked about coming up with test hypotheses based on conversion research.

Now we need to test the issues to validate and learn. Pick a testing tool, and create treatments / alternative variations to test against the current page (control).

There's no shortage of testing tools, one even built into Google Analytics and completely free. I use Optimizely mostly, but there's also VWO, Qubit, Adobe Target, Convert.com, and many others.

A thing to keep in mind is that you want to take testing seriously, you either need the help of a developer, or you need to learn some HTML, CSS, and javascript/jquery.

Testing is no joke – you have to test right. Bad testing is even worse than no testing at all – the reason is that you might be confident that solutions A, B, and C work well while in reality they hurt your business. Poor A/B testing methodologies are costing online retailers up to $13bn a year in lost revenue, according to research from Qubit. Don't take this lightly!

Very typical story of a business that does a/b testing is that they run 100 tests over the year, yet a year later their conversion rate is where it was when they began. Why? Because they did it wrong. Massive waste of time, money and human potential.

You need to make sure your sample size is big enough

In order to be confident that the results of your test are actually valid, you need to know how big of a sample size you need.

There are several calculators out there for this – like this or this.

You need a minimum number of observations for the right statistical power. Using the number you get from the sample size calculators as a ballpark is perfectly valid, but the test may not be as powerful as you had originally planned. The only real danger is in stopping the test early after looking at preliminary results. There's no penalty to have a larger sample size (only takes more time).

As a very rough ballpark I typically recommend ignoring your test results until you have at least 350 conversions per variation (or more – depending on the needed sample size). But don't make the mistake of thinking 350 is a magic number, it's not. Always calculate the needed sample size ahead of time! Related reading: Stopping A/B Tests: How Many Conversions Do I Need?

If you want to analyze your test results across segments, you need even more conversions. It's a good idea to run tests targeting a specific segment, e.g. you have separate tests for desktop, tablets and mobile.

Once your test has enough sample size, we want to see if one or more variations is better than Control. For this we look at statistical significance.

Statistical significance (also called statistical confidence) is the probability that a test result is accurate and not due to just chance alone. Noah from 37Signals said it well:

> *"Running an A/B test without thinking about statistical confidence is worse than not running a test at all—it gives you false confidence that you know what works for your site, when the truth is that you don't know any better than if you hadn't run the test."*
>
> *Most researchers use the 95% confidence level before making any conclusions. At 95% confidence level the likelihood of the result being random is very small. Basically we're saying "this change is not a fluke or caused by chance, it probably happened due to the changes we made".*

When an A/B testing dashboard (in Optimizely or a similar tool) says there is a "95% chance of beating original", it's asking the following question: Assuming there is no underlying difference between A and B, how often will we see a difference like we do in the data just by chance? The answer to that question is called the significance level, and "statistically significant results" mean that the significance level is low, e.g. 5% or 1%. Dashboards usually take the complement of this (e.g. 95% or 99%) and report it as a "chance of beating the original" or something like that.

If the results are not statistically significant, the results might be caused by random factors and there's no relationship between the changes you made and the test results (this called the null hypothesis).

But don't confuse statistical significance with validity.

Once your testing tool says you've achieved 95% statistical significance (or higher), that doesn't mean anything if you don't have enough sample size. Achieving significance is not a stopping rule for a test.

Read this blog post to learn why. It's very, very important.

Duration of the test

For some high-traffic sites you would get the needed sample size in a day or two. But that is not a representative sample – it does not include a full business cycle, all week days, weekends, phase of the moon, various traffic sources, your blog publishing and email newsletter schedule and all other variables.

So for a valid test both conditions – adequate sample size + long enough time to include all factors (full business cycle, better yet 2 cycles) – should be met. For most businesses this is 2 to 4 weeks.

No substitution for experience

Start running tests now.

There's quite a bit to know about all this, but the content above will already make you smarter about running tests than most marketers.

Read Next Lesson or Download guide as PDF

- #1: Mindset of an Optimizer

You seek to understand your customers better - their needs, sources of hesitation, conversations going on inside their minds.

- #2: Conversion Research

  Would you rather have a doctor operate on you based on an opinion, or careful examination and tests? Exactly. That's why we need to conduct proper conversion research.

- #3: Google Analytics for Conversion Optimization

  Where are the problems? What are the problems? How big are those problems? We can find answers in Google Analytics.

- #4: Mouse Tracking and Heat Maps

  We can record what people do with their mouse / trackpad, and can quantify that information. Some of that data is insightful.

- #5: Learning From Customers (Qualitative Surveys)

  When quantitative stuff tells you what, where and how much, then qualitative tells you 'why'. It often offers much more insight than anything else for coming up with winning test hypotheses.

- #6: Using Qualitative On-Site Surveys

  What's keeping people from taking action on your website? We can figure it out.

- #7: User Testing

  Your website is complicated and the copy doesn't make any sense to your customers. That's what user testing can tell you - along with specifics.

- #8: From Data to Test Hypotheses

  The success of your testing program depends on testing the right stuff. Here's how.

- #9: Getting A/B Testing Right

  Most A/B test run are meaningless - since people don't know how to run tests. You need to understand some basic math and statistical concepts. And you DON'T stop a test once it reaches significance.

- #10: Learning from Test Results

  So B was better than A. Now what? Or maybe the test ended in "no difference". But what about the insights hidden in segments? There's a ton of stuff to learn from test outcomes.

- Conclusion

  Conversion optimization is not a list of tactics. Either you have a process, or you don't know what you're doing.

COMPANY

About CXL
User Reviews
Pricing
Privacy Policy
Cookie Policy
Terms & Conditions
Service Level Agreement

RESOURCES

All courses
Free lessons and courses
Webinars
Instructors
Blog
Newsletter
A/B Test Calculator

TOP CATEGORIES

Demand Generation
Marketing Analytics
Brand Marketing
Conversion Rate Optimization
Growth Marketing
Product Marketing
SEO
PPC

TOP PROGRAMS

B2B Marketing
AI in Marketing
Conversion Rate Optimization
Product Marketing
Growth Marketing
Digital Analytics
Brand Marketing

LIVE COURSES

Unique positioning for agencies
Increase visibility and revenue from AI discovery engines
Run efficient organic and paid B2B campaigns
Optimize pages for AI Search with AEO
B2B Sales-marketing alignment
Competitive research for ads using AI
ICP modelling and targeting with AI

CONTACT

E-mail support@cxl.com

©2011-2025 CXL